Review for: ***"Evaluation of the New York State Mesonet Profiler Network Data"*** (amt-2022-85) submitted to *Atmospheric Measurement Techniques* by *Bhupal Shrestha et al.*

General recommendation:

Major revision

Synopsis:

The manuscript presents the observations by a network of microwave radiometer, Doppler lidar and all-sky camera in New York state. For three of the 17 stations, comparisons with nearby radiosondes have been performed and are analysed in terms of profiles of temperature, humidity and wind as well as for different thermodynamic indices that were derived from the original profiles. Furthermore, the aerosol optical depth from the camera observations is evaluated using AERONET sun photometer observations.

General comments:

The presentation of this quite dense network of ground-based remote sensing observations will be very useful to many readers if some more information is given on the quality of the dataset, especially regarding the microwave radiometer (MWR) dataset.

The Doppler lidar observations seem to agree well with the radiosondes. Sections 4.1 and 4.2 are well presented and can be published like this.

However, I am concerned for the large biases in the MWR observations compared to the radiosondes (especially for temperature). This cannot be only explained by retrieval uncertainties, especially the 8 K bias for clear sky cases. The reason for these large biases must be further investigated before the manuscript is ready for publication.

It is well known from literature that the quality of MWR temperature profiles decreases significantly with height. Therefore, it might be useful to limit the evaluation of the MWR profiles to the lower troposphere (e.g. 3 km), as for wind profiles.

Some more detailed questions and comments to the retrievals and their application:

- How were the neural network retrievals trained that you use for the MWR profiles? You are presenting the retrievals in lines 205ff., but I am wondering whether a retrieval self-test (by applying to forward-modelled radiosonde profiles) would result in similar biases. Do you have any information about that? If the self-test produces also the observed biases then there is an inherent problem with the retrieval (e.g. the setup of the radiative transfer model, the gas absorption models used, or the neural network design). On the other hand, if the self-test is bias-free then the reasons for the bias must stem from the instrument (e.g. wrong channel bandpasses, faulty calibration, etc.). However, the second hypothesis is much less likely, because in this case, the performance would not show such similar bias/error patterns at the different locations.
- Do you apply separate retrievals for 90° and 20° elevation observations? Or do you have any combined retrieval which uses brightness temperatures at both angles as input? How do the zenith and off-zenith observations compare e.g. during clear sky? I suppose that there is a stronger bias in the off-zenith observations as the optical path through the atmosphere is

much longer and there is less contribution to the signal from the higher levels due to attenuation (esp. for temperature profiles).

- The problem due to drifting radiosondes is rather minor and should not affect the retrieval bias (only the RMSE) as the sondes might randomly drift to warmer/colder/drier/moister regions. (see also p.19, l. 405-406).
- Concerning the calibration of the instruments: How often were the radiometers calibrated using liquid nitrogen? Do you see jumps in the performance before and after calibrations?
- I am a bit concerned by the bias correction. I am sure that on average you are improving the profiles, but how do you know that e.g. in severe weather situations (which are often at the edge of the statistical distribution), the bias correction performs well?
- For the bias correction, you are separating your dataset in clear sky/cloudy/rainy cases (Fig. 10a,b,c). What is the variability of these biases as a function of height for each of the three weather classes? Are there any seasonal differences? What about very cold winter days or hot summer conditions? I am pretty sure that the bias is quite variable, and the mean bias of 0°C / 0 gm^{-3} in Fig. 11 (a) and (d) does not show the range or the uncertainty of possible biases. Also, I would like to know more about the physical reason of the different biases depending on the weather.
- Usually, a bias correction is done before applying the retrieval, i.e. a correction of the brightness temperatures (as it is done e.g. in Lohnert and Maier, 2012), but I guess that this is beyond the scope of this paper.

Detailed comments:

- p.1, l.20-21: Why do you only mention biases here?
- p.3, l.81: The RMSE of <= 7 K is a very high number. Other publications show quite low numbers.
- p.6, l.140-141: The vertical resolution is determined by the retrieval algorithm. The "true" vertical resolution is very coarse and decreasing with height (see e.g. Crewell and Lohnert, 2007). For these passive observations, often the degrees of freedom are determined, which is between 2 and 5 for the whole profile (depending on the choice of frequencies and angles).
- p.6, table 1: Are you sure that the manufacturer gives a relative humidity accuracy of 2% ?
- p.16, fig. 6 a-c: These scatterplots are not very useful, as the performance of the MWR retrievals depends strongly on the height above ground (and not on the temperature itself). You could try to make a color plot Temperature bias vs. height for bins of 1°C vs. 100m (or the vertical resolution of the retrieval)
- p.19, l.396-397: Yes, I can see that the biases are significant, but you should further investigate the reason for that.
- p. 19, l. 398ff.: *"This is likely because the MWR measured V-band observations are ingested into the neural network with greater weighting function at lower heights than at higher heights to produce a finer vertical resolution at lower heights."*
  This phrase makes no sense to me. The neural network itself will determine the weights of the single V-band brightness temperature observations for the temperature profile. This has nothing to do with the vertical resolution.
- p.19, l.401-402: *"In contrast, the water vapor density comparisons show better results at higher altitudes than at lower altitude"*

This is not surprising, as there is not much water vapor (absolute) in upper levels. If you compared the relative error, the result would look very different.

- p.20, table 4: This table would be more interesting as a function of height (or limited to the lower troposphere, such as up to 3 km)